For the virtual issue, "Red Sea and Western Indian Ocean Biogeography"

Original Article

LRH: I. Fernandez-Silva *et al*.

RRH: Phylogeography of a tropical goatfish

Yellow tails in a Red Sea: Phylogeography of the Indo-Pacific goatfish *Mulloidichthys flavolineatus* reveals isolation in peripheral provinces and cryptic evolutionary lineages

Iria Fernandez-Silva[1,2*], John E. Randall[3], Richard R. Coleman[1], Joseph D. DiBattista[4,5], Luiz A. Rocha[2], James D. Reimer[6], Carl G. Meyer[1], Brian W. Bowen[1]

[1] *Hawai'i Institute of Marine Biology, University of Hawai'i, Kāne'ohe, HI 96744, USA,* [2]*Section of Ichthyology, California Academy of Sciences, San Francisco, CA 94118, USA,* [3]*Bishop Museum, Honolulu, HI 96817, USA,* [4]*Red Sea Research Center, Division of Biological and Environmental Science and Engineering, King Abdullah University of Science and Technology, Thuwal, 23955, Saudi Arabia,* [5]*Department of Environment and Agriculture, Curtin University, PO Box U1987, Perth, WA 6845, Australia,* [6]*Molecular Invertebrate Systematics and Ecology Laboratory, Graduate School of Engineering and Science, University of the Ryukyus, Okinawa 903-0213, Japan*

Correspondence: Iria Fernandez-Silva, Section of Ichthyology, California Academy of Sciences, 55 Music Concourse Drive, San Francisco, CA 94118, USA.

E-mail: iriafernandezsilva@gmail.com

Word count: 6600; Estimate of the number of pages for figures and tables: 2.5

ABSTRACT

Aim Broadly distributed reef fishes tend to have high gene flow mediated by a pelagic larval phase. Here we survey a reef-associated fish distributed across half the tropical oceans, from the Red Sea to the central Pacific. Our goal is to determine whether genetic structure of the broadly distributed Yellowstripe Goatfish (*Mulloidichthys flavolineatus*) is defined by biogeographic barriers, or smoothed via larval dispersal.

Location Red Sea, Indian Ocean, Pacific Ocean

Methods Specimens were obtained at 19 locations from the Red Sea to Hawai‘i. Genetic data include mtDNA cytochrome *b* (N = 217) and twelve microsatellite loci (N = 185). Analysis of molecular variance (AMOVA), STRUCTURE, a parsimony network and coalescence analyses were used to resolve recent population history and connectivity.

Results Population structure was significant (mtDNA ϕ_{ST} = 0.68, P < 0.001; microsatellite F_{ST} = 0.08, P < 0.001), but mostly driven by samples from the North-western (NW) Indian Ocean (including the Red Sea) and Hawai‘i. There was little population structure across the Indian Ocean to the central Pacific. Hawai‘i was distinguished as a highly isolated population (mtDNA ϕ_{ST} = 0.03 to 0.08, P = N. S.; microsatellite F_{ST} = 0.05 to 0.10, P < 0.001). Specimens from the NW Indian Ocean clustered as a distinct phylogenetic lineage that diverged approximately 493

ka (d =1.7%), which indicates that these fish persisted in isolation through several Pleistocene glacial cycles.

Main Conclusion These data reinforce the emerging themes that: 1) phylogeographical breaks within species often coincide with biogeographical breaks based on species distributions and 2) populations on the periphery of the range (NW Indian Ocean and Hawai‘i) are isolated and these areas may be evolutionary incubators.

Keywords

Cytochrome *b*, fish, glacial refugia, Hawai‘i, marine biogeography, microsatellites, mtDNA, population structure, Red Sea, vicariance

INTRODUCTION

Broadly distributed reef fishes tend to show high connectivity and gene flow mediated by a pelagic larval phase that can last from weeks to months. However, geographical and oceanographical features, as well as ecological gradients, can act as barriers to gene flow, and as such they often set the boundaries of biogeographical provinces. Their persistence through tectonic movements and glacial cycles has an enormous influence on the evolution of marine life (Rocha *et al*., 2007).

The Indo-Pacific is the largest continuous biogeographic province on Earth. Here, the tropical marine provinces are characterized by many wide-ranging species, and it is estimated that more than 1500 fish species have distributional ranges extending more than 10,000,000 km^2 (Allen, 2008; Briggs & Bowen, 2012). Over such a vast geographical scale, the spatial

distribution of genetic variation is particularly useful to place the bounds over which populations are likely to be demographically linked (Lowe & Allendorf, 2010) or evolutionarily independent (Rocha *et al*., 2007). For example, genetic studies have revealed cryptic lineages within a single nominal species that led to subsequent recognition of sister taxa (DiBattista *et al*., 2011). In other cases, species recognized on the basis of distinct morphology or colouration show no genetic differences (DiBattista *et al.*, 2012). The proper alignment of species boundaries is fundamental for effective management of marine resources, as well as to identify areas of elevated endemism and biodiversity hotspots.

Here we contribute a phylogeographical study of a tropical shore fish, the yellowstripe goatfish, *Mulloidichthys flavolineatus* (Lacepède, 1801) (Mullidae). *Mulloidichthys flavolineatus* is distributed from the Red Sea and East Africa to the Hawaiian, Marquesan and Pitcairn Islands, north to the Ryukyus and Bonin Islands and south to Montague Island in New South Wales (Randall, 2002). Where present, *M. flavolineatus* tends to be common and frequently observed in sandy areas around coral reefs (1 to 76 m; Randall, 2002). They are valued food and an important catch for many coastal communities. Despite its economical and ecological importance, a comprehensive assessment of intraspecific variability and dispersal in *M. flavolineatus* has not been conducted.

We obtained samples from almost the entire distribution of *M. flavolineatus* and generated data from mitochondrial DNA (mtDNA) sequences and nuclear microsatellites to investigate phylogeographical patterns. In doing so, we discovered the presence of a cryptic lineage endemic to the Red Sea and Arabian Sea. The geographic distribution of this cryptic lineage follows the biogeographical delineation of the North-western (NW) Indian Ocean province (Kulbicki *et al*., 2013; but see Briggs & Bowen, 2012). We also demonstrate genetic

isolation of the Hawaiian population. With these results, we discuss the evolutionary mechanisms that promote elevated endemism in the NW Indian Ocean and Hawai‘i, and the role of peripheral provinces as sources of evolutionary innovation.

MATERIALS AND METHODS

Collections and molecular analyses

Genomic DNA was extracted following Meeker *et al*. (2007) from *M. flavolineatus* fin clips collected at 19 locations in the Red Sea, Arabian Sea, Indian Ocean and Pacific Ocean (Fig. 1, Table 1). A fragment of the mtDNA cytochrome *b* gene (*cyt b*) (Meyer, 1993) was sequenced as described in Appendix S1. We aligned, edited and trimmed the sequences to a common length using GENEIOUS PRO *vers.* 6.0 (Biomatters, LTD, Auckland, NZ). All haplotypes were identified and deposited in GenBank (accession numbers: XXX-XXX) and Dryad (http://www.datadryad.org/).

For locations with $N > 5$ we applied twelve previously described microsatellite loci (Fernandez-Silva *et al.*, 2013) as described in Appendix S1. Briefly, microsatellite loci were amplified by polymerase chain reaction (PCR) following the M13-tailed primer method (Fernandez-Silva *et al.*, 2010), and resolved using an ABI3130XL capillary sequencer (Applied Biosystems, Foster City, CA, USA). Due to logistic constraints we were unable to genotype the samples from the Gulf of Aqaba/Eilat ($N = 12$) with microsatellites.

Population genetic analyses

Ten populations (191 individuals) had samples sizes greater than five and were thus suitable for population genetic analyses; however we included all (217 individuals) for the construction of

115 the mtDNA haplotype network outlined below. ARLEQUIN *vers.* 3.5.1.3 (Excoffier & Lischer,
116 2010) was used to calculate mtDNA haplotype (h) (Nei, 1987) and nucleotide (π) diversity (Nei
117 & Li, 1979) and to investigate population structure at various geographical scales. To test for
118 genetic partitioning, we applied an analysis of molecular variance (AMOVA) based on Φ_{ST}, an
119 analogue of Wright's F_{ST} that incorporates a model of sequence evolution (Weir & Cockerham,
120 1984; Excoffier *et al*., 1992). Since the specimens from the NW Indian Ocean were
121 phylogenetically distinct, we used the following groupings: (1) within the Red Sea to Arabian
122 Sea, including Gulf of Aqaba/Eilat, Saudi Arabian Red Sea, Djibouti and Oman (hereafter
123 referred to as the NW Indian province), (2) within the group formed by all remaining sampling
124 sites in the Indian Ocean and Pacific Ocean (Madagascar, Ryukyus, Cook Islands, Line Islands,
125 Johnston Atoll and Hawai‘i [hereafter referred to as the Indo-Pacific region]), and (3) between
126 these two groups. We applied the K80 model (Kimura, 1980) allowing rate variation among sites
127 (Gamma parameter a = 0.07) as indicated by JMODELTEST2 (Guindon & Gascuel, 2003; Darriba
128 *et al*., 2012). We used non-parametric permutation procedures (10,000 iterations) to test the
129 significance of the variance components for each hierarchical comparison and applied the
130 modified false discovery rate method (Narum, 2006) to correct for multiple comparisons. We
131 also calculated Jost's D_{est}, an absolute measure of differentiation between sites that is not affected
132 by within-population diversity (Jost, 2008), with the program SMOGD *vers.* 1.2.5 (Crawford,
133 2010).

134 An unrooted network was constructed using Network *vers.* 4.6.0.0 and NETWORK
135 PUBLISHER *vers.* 1.3.0.0 (http://www.Fluxus-engineering.com/network_terms.htm), applying the
136 median-joining and maximum-parsimony options. We calculated corrected average divergence

between mitochondrial lineages (*d*) with MEGA6 (Tamura *et al*., 2013), assuming the K80 model and Gamma parameter $a = 0.07$.

We applied the Bayesian MCMC approach implemented in BEAST *vers.* 1.7.5 (Drummond *et al*., 2012) to estimate coalescent times in *M. flavolineatus*. The analysis was conducted under the HKY model of mutation (allowing for rate variation among sites) with a strict clock, using a rate of change of 0.65% per Myr, which is the average reported for six trans-isthmian geminate reef fishes that are assumed to have diverged as a consequence of the closure of the Isthmus of Panama 3.0 Ma (Lessios, 2008). We also repeated the analyses assuming a faster (1.0 % per Myr) molecular clock (Bowen *et al*., 2001). We computed three independent runs (10^7 generations, after 10^6 burn-in chains) and combined the continuous parameter values sampled every 10^3 generations using the program TRACER *vers.* 1.6 (http://tree.bio.ed.ac.uk/software/tracer/).

To assess deviations from neutral mutation-drift equilibrium, we estimated Fu's F_S and Tajima's *D* with ARLEQUIN (99,999 permutations). To further investigate the historical demography of *M. flavolineatus* we calculated mismatch distributions (Schneider & Excoffier, 1999) and Harpending's raggedness index (Rogers & Harpending, 1992). Two individuals sampled in Jeddah were excluded from the historical demography analyses because they belonged to the Indo-Pacific lineage rather than the NW Indian Ocean lineage (see Results).

Microsatellite diversity was calculated as observed (H_O) and expected (H_E) heterozygosity per locus and per sampling location using GENODIVE *vers.* 2.0 (Meirmans & Van Tienderen, 2004). We calculated the effective number of alleles as $Eff\_allele = 1/(1- H_E)$. Deviations from Hardy-Weinberg equilibrium (HWE) and linkage equilibrium were assessed with ARLEQUIN using default parameters. The fixation index F_{ST} (Excoffier *et al*., 1992) was

160 calculated with GENODIVE, along with standardized versions G'_{ST} (Nei, 1987) and G''_{ST} (Hedrick,
161 2005), that account for within-population diversity (H_S) to facilitate comparisons across loci. We
162 also calculated per locus values of Jost's $D_{est.}$ We applied the method in LOSITAN *vers.* 1.0.0
163 (Antao *et al.*, 2008) to identify outlier loci that have excessively high or low F_{ST} compared to
164 neutral expectations.

165 We analysed genetic structure with microsatellites assuming the infinite allele model
166 under the AMOVA framework using ARLEQUIN with 99,999 permutations, with F_{ST} equivalent to
167 Weir & Cockerham's θ (Weir & Cockerham, 1984). To ensure that observed patterns of genetic
168 structure (or lack thereof) were not driven by the effect of individual loci, we re-calculated
169 indexes of population differentiation after sequential removal of loci and compared the results.
170 Population genetic structure across sampling locations was further evaluated using a Bayesian
171 clustering analysis implemented in STRUCTURE *vers.* 2.3.4 (Pritchard *et al.*, 2000). This method
172 groups multi-locus genotypes to minimize deviations from Hardy Weinberg and linkage
173 equilibrium. We applied a model of admixture and correlated allele frequencies, and included the
174 geographical origin of genotypes as a prior. We determined the most likely number of clusters
175 (K) in the dataset by comparing the mean probability of the data (Ln P(D)) over five replicate
176 runs (400,000 iterations after a burn-in of 100,000 repetitions) for each value of K = 1 to K = 5.
177 The method described by Evanno *et al.* (2005) produced similar results.

178 To test for isolation-by-distance (IBD), we conducted Mantel tests using linearized F_{ST}
179 [$F_{ST}/(1 - F_{ST})$] and the natural log of Euclidean distance among all sample sites (Rousset, 1997)
180 in ARLEQUIN using the *cyt b* and microsatellite data sets independently. To account for the fact
181 that IBD can be confounded by hierarchical genetic structure (Meirmans, 2012), we tested IBD

separately in the NW Indian Ocean and remaining Indo-Pacific regions (excluding Johnston Atoll and Hawai‘i).

RESULTS

Molecular diversity: mtDNA

We resolved 715 bp of the *cyt b* gene from 217 individuals (Fig. 1 and Table 1). Across all sites, haplotype diversity (*h*) was high, ranging from 0.52 to 0.96, with a mean of 0.80 (when populations with more than five individuals were considered; Table 1). In contrast, nucleotide diversity (π) was relatively low, ranging from 0.002 to 0.007, with a mean of 0.004. The haplotype network (Fig. 2) revealed a genetic and geographical partition of the goatfish into two clusters, one composed exclusively of haplotypes from the NW Indian Ocean province and another cluster composed of all the haplotypes from the remaining Indo-Pacific region plus two unique haplotypes from Jeddah (Saudi Arabia). The NW Indian sample ($N = 59$) included 17 polymorphic sites defining 16 haplotypes. The most common haplotype (39%) was detected at every site. In the remaining Indo-Pacific ($N = 158$), we detected 40 polymorphic sites defining 35 haplotypes, and the most common haplotype (46%) was observed in all sites with $N > 2$. Although in this region a great proportion of haplotypes are shared across sites, shifts in haplotype frequencies indicate isolation of the Hawaiian province (Hawai‘i and Johnston Atoll).

Pairwise mismatch distributions did not show significant deviation from the simulated demographic expansion null model in any of the regions tested (Table 2), although Harpending’s raggedness index was marginally significant in the Indo-Pacific region when the samples from the Hawaiian province were excluded ($r = 0.086$, $P = 0.04$). Significant values of Fu’s *F*s and Tajima’s *D* support a scenario of past demographic expansion across all regions (Fu’s *Fs*= -28.55

to -6.90, $P < 0.01$; Tajima's D = -1.24 to -2.32, $P < 0.09$). Both the NW Indian and Indo-Pacific clades had starburst patterns in the haplotype networks (Fig. 1), supporting the conclusion of recent population expansion. We verified that the pattern held when we constructed separate haplotype networks with the Hawaiian province and remainder of the Indo-Pacific (results not shown).

The corrected average pairwise sequence divergence between clades was d = 1.7% (Kimura, 1980), much higher than the genetic divergence within each cluster (NW Indian: 0.4%; Indo-Pacific: 0.3 %). Assuming a 0.65% per Myr, coalescence time for *M. flavolineatus* was approximately 494 ka [95% highest posterior density intervals (HPD) of 307 and 706 ka]. Coalescence time was 313 ka (95% HPD = 192 to 447 ka) when assuming a faster (1.0% per Myr) molecular clock. Coalescence times of the NW Indian and Indo-Pacific lineages were estimated at approximately 265 ka (95% HPD = 154 to 391 ka) and 298 ka (95% HPD = 187 to 434 ka), respectively, assuming 0.65% per Myr.

Molecular diversity: Microsatellites

We genotyped 185 individuals at twelve microsatellite loci. Number of alleles ranged from 4.7 to 6.7 (multi-locus average) across the nine populations surveyed (Table 1). Expected heterozygosity (H_e) ranged from 0.58 to 0.73, while observed heterozygosity (H_o) exhibited a slightly narrower range, from 0.53 to 0.68 (Table 1). After applying a false discovery rate correction, we found that only one of 108 tests of Hardy-Weinberg equilibrium (Hawai‘i, locus mflssr034, $P < 0.001$) and 8 of 594 tests for linkage disequilibrium (with biases towards mflssr026 and mflssr033) were significant at 95% confidence.

Two loci had comparatively higher F'_{ST} and G'_{ST} values than the multi-locus average (mflssr034 and mflssr037, Table S1 in Appendix S2), indicating the presence of geographically restricted alleles, possibly owing to positive directional selection acting upon these or linked loci.

Population structure

We detected a sharp genetic subdivision between the NW Indian populations and the remainder of Indo-Pacific as indicated by the AMOVA analyses. This break explained most of the variability in mtDNA (Φ_{CT} = 0.91, P = 0.01); the variance explained by among-populations-within-regions was negligible in comparison (Φ_{SC} = -0.0004, P = 0.56). Whereas AMOVA tests with microsatellite data also support this break (F_{CT} = 0.10, P = 0.01), the degree of genetic variation among populations within regions was still high (F_{SC} = 0.05, P < 0.001), indicating additional genetic sub-structuring.

Consistently, all pairwise Φ_{ST} or F_{ST} values in population comparisons between NW Indian and Indo-Pacific regions were significant and much higher than within regions (Table 3). Standardized F'_{ST} and G'_{ST} and true population differentiation (D_{est}) indexes followed the same pattern (Tables S2 and S3 in Appendix S2). Locus mflssr034 showed marked segregation of alleles between the NW Indian province and Indo-Pacific regions (F_{ST} = 0.26 to 0.56, mean F_{ST} = 0.42; within the Indo-Pacific: F_{ST} = 0 to 0.01, mean F_{ST} = 0.03; within the North-western Indian province: F_{ST} = 0, Table S4 in Appendix S2).

Within each geographical region, population pairwise tests provide insights into genetic sub-structuring when differentiation is present, subtle or absent (Table 3). Samples from Hawai‘i and Johnston Atoll showed genetic differentiation from all other Indo-Pacific locations (microsatellites: F_{ST} = 0.04 to 0.10, P < 0.001; mtDNA: pairwise Φ_{ST} = 0 to 0.10, N.S). In comparisons among the Line Islands, the Cook Islands and Madagascar, F_{ST} values were

significant in four out of six tests with microsatellite loci only, but no comparisons were significant when locus mflssr037 was removed from the analyses (Table 3 and Table S5 in Appendix S2). We found no significant genetic differentiation in comparisons within the NW Indian province or between Hawai‘i and Johnston Atoll. Results from the various AMOVA analyses were consistent with these findings (Table 4). However, we cannot exclude the possibility that increasing sample size in the NW Indian province may enhance statistical power and detect genetic differentiation. Tests for IBD were not significant for *M. flavolineatus.*

Bayesian clustering analyses implemented in STRUCTURE identified the mean estimated probability of the data as being highest when samples are grouped into three clusters ($K = 3$) (Fig. 3 and Fig. S1 in Appendix S2). All individuals from Hawai‘i and Johnson Atoll were assigned to Bayesian Cluster 1 (BC1), all individuals from the Line Islands, the Ryukyus and Cook Islands were assigned to BC2, and all individuals from Djibouti, Oman and Jeddah were assigned to BC3. The sample from Madagascar was predominantly assigned to BC2, but three individuals were assigned to BC3. Notably, the two individuals from Jeddah with mtDNA haplotypes characteristic of the Indo-Pacific *cyt b* clade (see Fig. 2) grouped with individuals from the NW Indian province in BC3 based on microsatellite genotypes. In most cases individuals were assigned with high confidence to one of three clusters (>95% assigned with >80% probability), with a few exceptions (Cook Islands).

DISCUSSION

The goatfish *M. flavolineatus* shows significant levels of genetic structure that indicate an evolutionary partition between the NW Indian province and the greater Indo-Pacific regions. We

also detected a significant, but more recent, population genetic partition between Hawai‘i and our sample locations in the Indian and Pacific Oceans (Tables 3 & 4 and Fig. 1, 2 & 3).

It is remarkable that the population from the NW Indian province maintained separation from Indian Ocean counterparts despite contemporary connection between these water masses (Nuryanto & Kochzius, 2009; DiBattista *et al*., 2013). The isolation of Red Sea marine fauna is well documented, with many endemic reef fishes (13%), annelids (13%), arthropods (10%), tunicates (17%), echinoderms (8%), molluscs (6%), and corals (6%) (DiBattista *et al*., in review). Several evolutionary processes may favor this isolation. Sea-level lowering during Pleistocene glacial periods affected the exchange of water through the narrow (18 km) and shallow (137 m) Straight of Bab al Mandab (Siddall *et al*., 2003), which may have acted as a barrier to larval transport. Reduction of water exchange led to increased residence times of Red Sea water, which resulted in salinity levels up to 50 psu and harsh conditions for survival (Fenton *et al.,* 2000). In addition, cold-water upwelling off the northeast African and southern Arabian coasts, together with a 2200 km stretch of coast between Southern Somalia and Djibouti that lacks true coral reefs, may reinforce the isolation of the NW Indian province from the greater Indian Ocean (Kemp, 1998).

The two mitochondrial lineages (NW Indian and Indo-Pacific) became isolated from each other approximately 493 ka based on a molecular clock of 0.65% per My (Lessios, 2008). These lineages remained highly isolated as indicated by the absence of admixture detected with microsatellite markers. The timing of divergence precedes the most recent desiccation event in the Red Sea (~19 ka), as well as most salinity crises of the Pleistocene (Siddall *et al.*, 2003). This timing indicates that NW Indian populations of *M. flavolineatus* did not go extinct with each ice age, a finding that reinforces the emerging view that some NW Indian populations of reef fauna

294 survived multiple salinity crises (DiBattista *et al.*, 2013; Hodge *et al.*, 2014).

295 Whether the North-western Indian lineage persisted just outside of the Red Sea or refugia
296 existed within the Red Sea is debatable (DiBattista *et al.*, in review 2). Persistence in refugia
297 outside of the Red Sea in the Gulf of Aden (Djibouti or mainland Yemen) is compatible with our
298 results, given that the cold-water upwelling off the northeast African and southern Arabian coasts
299 may account for the isolation of the NW Indian and Indo-Pacific clades, although the historical
300 continuity of this upwelling is poorly characterized (Rampen *et al.*, 2008). Alternatively,
301 persistence through Pleistocene glacial ages in the Red Sea and later colonization of the Gulf of
302 Aden and Omani coast is feasible. Support for Red Sea refugia comes from fishes endemic to the
303 northern Red Sea, for example, *Sillago suezensis* (Sillaginidae), *Atherinomorus forskalii*
304 (Atherinidae), *Upeneus davidaromi* (Mullidae) and *Diplogrammus gruveli* (Callionymidae)
305 (Golani *et al.* 2001; Tikochinski *et al*., 2013; Randall *et al*., in press). It may be that species that
306 persisted in the Red Sea during these harsh periods have broader habitat requirements, and
307 therefore are less reliant on coral reefs. In considering habitat during low sea level stands, *M.*
308 *flavolineatus* prefers sandy, shallow substrates, including lagoons without coral reefs. Thus, a
309 third possibility is that the Yellowstripe Goatfish persisted in the sediment-loaded southern Red
310 Sea and that its habitat was connected with the Gulf of Aden during low sea level stands.
311 However, given that survival of reef fauna in the Red Sea during glaciations is contentious, and
312 lack of genetic differentiation between the Red Sea and NW Indian populations, the most
313 parsimonious explanation is that *M. flavolineatus* persisted outside the Red Sea rather than
314 within the Red Sea.

315 Locus mflssr034 (or loci linked to it) shows strong differentiation between the NW
316 Indian province and greater Indo-Pacific regions, a genomic signature that invokes the possibility

317 that selection and local adaptation played a role in lineage divergence (Willette *et al.* 2014).
318 Given the history of dramatic changes in the physical and biotic environment in the region, the
319 possibility of divergence along ecological boundaries is compelling (Rocha *et al*., 2005). Hence,
320 we favour an evolutionary scenario of parapatric speciation, where vicariance due to
321 biogeographic barriers is reinforced by divergent adaptation in response to ecological
322 differences.

323 **Taxonomic considerations and endemism in the Red Sea**

324 MtDNA divergence between NW Indian and Indo-Pacific *M. flavolineatus* is higher than the ge-
325 netic distance among four of the other seven species pairs in the genus (*M. vanicolensis, M. den-*
326 *tatus, M. martinicus* and *M. mimicus*) (Results not shown). Furthermore, high levels of mito-
327 chondrial and nuclear differentiation are concordant with differences in morphology (Appendix
328 3) and coloration of the caudal fin (Fig. 2). In light of these observations, we are currently con-
329 ducting additional taxonomic work to determine whether a revision of *M. flavolineatus* is appro-
330 priate (I. Fernandez-Silva *et al., in prep*).

331 The discovery of an endemic goatfish lineage in the NW Indian province is not
332 surprising. Among 84 described species of goatfishes (family Mullidae), seventeen are found in
333 the Red Sea, four of which are endemic (Randall *et al*., in press): *Upeneus pori* Ben-Tuvia &
334 Golani, 1989, *U. davidaromi* Golani, 2001, *U. niebuhri* Guézé, 1979 and *Parupeneus forskali*
335 (Fourmanoir & Guézé, 1976). Our study exemplifies how phylogeographical analyses provide a
336 framework for uncovering cryptic evolutionary lineages with relevance to marine conservation.
337 Because there has been little genetic work on widespread taxa that range across the Red Sea and
338 other biogeographical provinces (reviewed by Berumen *et al.*, 2013), endemism estimates for the
339 Red Sea (or NW Indian province) are likely to increase (DiBattista *et al*., in review).

The isolation of Hawai‘i

For the Indo-Pacific lineage, our combined analyses of mtDNA and microsatellite markers revealed population level differentiation between *M. flavolineatus* in Hawai‘i and the remainder of the Indo-Pacific (Table 3 & 4, and Fig. 1 & 3). Hawai‘i is the most isolated province on the Pacific plate (Hourigan & Reese, 1987; Randall, 2007). 25% of Hawaiian fishes, 25 % of the red algae and 20% of the molluscs are endemic (Briggs & Bowen, 2012, and references therein), and many studies have documented the genetic isolation of Hawaiian populations (Planes & Fauvelot, 2002; Gaither *et al*., 2011). The coalescence time of the Indo-Pacific populations (~287 ka) indicates that the separation of the Hawaiian province was more recent than the split between the NW Indian and Indo-Pacific lineages (~497 ka). The isolation of the Hawaiian populations coincides with the timing of divergence of many Hawaiian endemics (<1 Ma, Hodge *et al.*, 2014).

Our study indicates genetic alignment of Johnston Atoll with Hawai‘i, congruent with previous work that has documented a high degree of genetic connectivity between these two locations (DiBattista *et al*., 2011), with a strong component of Hawaiian endemics in the fish assemblages of Johnston Atoll (Gosline, 1955). Furthermore, simulations of larval transport from Johnston Atoll to Hawai‘i have shown potential connectivity between the two regions (Kobayashi, 2006). Together with a genetic signature of population expansion in this province (Table 2), our results indicate that the Hawaiian populations originated from rare dispersal events from the South Pacific (possibly via Johnston Atoll). An alternative colonization route from the West Pacific into Hawai‘i via the warm Kuroshio Current is also possible (Pyle, 1999; Bird *et al*., 2011).

Outside of Hawai‘i, the Indo-Pacific lineage of *M. flavolineatus* demonstrates no obvious

363 geographical patterns of genetic differentiation. We also do not see an Indian Ocean versus
364 Pacific Ocean split (Table 3 & 4 and Fig. 3) as observed in other widespread taxa (Gaither &
365 Rocha, 2013). Mitochondrial haplotypes are shared among localities separated by more than
366 15,000 km (Fig. 1 & 2), demonstrating the high dispersal capabilities of *M. flavolineatus*. This
367 dispersal capability is also evident in its wide distribution – *M. flavolineatus* has colonized
368 throughout the Indian and Pacific oceans, except the Tropical Eastern Pacific and the
369 Arabian/Persian Gulf. An extended pelagic larval duration of 60 days (Luiz *et al*., 2013) and
370 schooling behaviour from larval stages (Leis & Carson-Ewart, 1998) to adulthood are some of
371 the life history traits that may enhance the capacity of *M. flavolineatus* to disperse and establish
372 new populations (Leis, 2006; Luiz *et al.,* 2013) .

373 **Peripheral populations as sources of evolutionary innovation**

374 Although previous work found no genetic structure in *M. flavolineatus* (Lessios & Robertson,
375 2013), by including a larger sample size, broader geographical sampling and more molecular
376 markers, our study revealed isolation of peripheral populations in the NW Indian Ocean and
377 Hawai‘i, and the genetic continuity of populations in the core of the Indo-Pacific. These results
378 are relevant to the long-standing debate over whether the Indo-Pacific biodiversity hotspot (i.e.
379 the Coral Triangle) is a \centre of \speciation versus a centre of accumulation and/or overlap
380 (Bowen *et al*., 2013). With this study we contribute another example (Nuryanto & Kochzius,
381 2009; DiBattista *et al*., 2011, 2013; Hodge *et al*., 2012; Tikochinski *et al*., 2013) that illustrates
382 how coral reefs at the periphery of the Indo-Pacific are sources of evolutionary innovation,
383 contributing to global marine biodiversity.

384

ACKNOWLEDGEMENTS

Many thanks to F. Ramahatratra, D. Aube, S. Samejima, D. Golani and J. Mclaughlin for the generous donation of entire sample sets from Madagascar, the Big Island, Yoron, Gulf of Eilat/Aqaba and Palmyra, respectively. For specimen collection we are grateful to T. Alpermann, G. Bernardi, C. Braun, J. Copus, M. Craig, R. Diaz, J. Eble, T. Fujii, M. Gaither, G. Golicha, R. Kosaki, J.P. Hobbs, Y. Irei, K. Koeda, T. Lilley, G. Nanninga, Y. Papastamatiou, M. Priest, M. Ramsey, J. Reece, P. Saenz-Agudelo, S. Hiasashi, T. Sinclair-Taylor, D. Smith, B. Taniguchi, K. Tenggardjaja, E. Tong, Yamada-san, D. Uyeno, the crew of the Whitness' Eyes and Hi'ialakai. For valuable advice and logistic support, our most sincere thanks to J.-A. Leong, R. Toonen, R. Kosaki, M. Berumen, H. Randall, L. O'Hara, S. Nakachi, F. Ublein, S. Bogorodsky, C. Sheppard, S. Karl, Z. Forsman, K. Andrews, C. Bird, M. Iacchei, D. Pence, E. Jessup, J. McIlwain, O. Masami, T. Naruse, O. Takama, E. Kawai, E. Mason, N. Prévot, members and alumni of the ToBo Lab at the Hawai'i Institute of Marine Biology, the Department of Ichthyology at California Academy of Sciences, MISE and Tachihara's Lab at Ryudai, and the Marine Station at Kuroshio, the Coral Reef Ecology Lab at KAUST and members of CMOR including A. Gusti. The authors thank S. Hou, A. Young (ASGPB) and A. Eggers (HIMB EPSCoR Core Genetics Facility) for their assistance with sequencing and genotyping. We also thank Papahānaumokuākea Marine National Monument, Hawai'i DLNR, USFWS, TNC (Hawai'i), WCGA, the Red Sea State Government and The Red Sea University, as well as Equipe Cousteau including N. Hussey, S. Kessel, C. Scarpellini and M. Younis; the Administration of the British Indian Ocean Territory, the Ministry of Agriculture and Fisheries in Oman including A. Karim; the Ministère de la Pêche et des Résources Halieutiques – Madagascar, as well as the Univ. Milano-Bicocca MaRHE Centre in Magoodhoo, the Ministry of

Fisheries and Agriculture, Maldives and the community of Maghoodhoo, Faafu Atoll. This research was supported by a National Science Foundation grant OCE-0929031 to B.W.B., the KAUST Office of Competitive Research Funds (OCRF) under Award No. CRG-1-2012-BER-002 and baseline research funds to M.L.B., as well as a Fulbright and Marie Curie (ERC) fellowships to I.F.S. and a National Geographic Society Grant 9024-11 to J.D.D. This is contribution no. XXX from the Hawai‘i Institute of Marine Biology and no. XXX from the University of Hawai‘i School of Ocean and Earth Science and Technology.

REFERENCES

Allen G.R. (2008) Conservation hotspots of biodiversity and endemism for Indo- Pacific coral reef fishes. *Aquatic Conservation: Marine and Freshwater Ecosystems*, **18**, 541–556.

Antao T., Lopes A., Lopes R.J., Beja-Pereira A., & Luikart G. (2008) LOSITAN: a workbench to detect molecular adaptation based on a Fst-outlier method. *BMC Bioinformatics*, **9**, 323.

Berumen M.L., Hoey A.S., Bass W.H., Bouwmeester J., Catania D., Cochran J.E.M., Khalil M.T., Miyake S., Mughal M.R., Spaet J.L.Y., & Saenz-Agudelo P. (2013) The status of coral reef ecology research in the Red Sea. *Coral Reefs*, **32**, 737–748.

Bird C.E., Holland B.S., Bowen B.W., & Toonen R.J. (2011) Diversification of sympatric broadcast-spawning limpets (*Cellana* spp.) within the Hawaiian archipelago. *Molecular Ecology*, **20**, 2128–2141.

Bowen B., Bass A., & Rocha L. (2001) Phylogeography of the trumpetfishes (*Aulostomus*): ring species complex on a global scale. *Evolution*, **55**, 1029–1039.

Bowen B.W., Rocha L.A., Toonen R.J., & Karl S.A (2013) The origins of tropical marine biodiversity. *Trends in Ecology & Evolution*, **28**, 359–66.

Briggs J.C. & Bowen B.W. (2012) A realignment of marine biogeographic provinces with particular reference to fish distributions. *Journal of Biogeography*, **39**, 12–30.

Crawford N.G. (2010) Smogd: Software for the measurement of genetic diversity. *Molecular Ecology Resources*, **10**, 556–557.

Darriba D., Taboada G., Doallo R., & Posada D. (2012) jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods*, **9**, 772–772.

438 DiBattista J.D., Berumen M.L., Gaither M.R., Rocha L.A., Eble J.A., Choat J.H., Craig M.T.,
439 Skillings D.J., & Bowen B.W. (2013) After continents divide: comparative phylogeography
440 of reef fishes from the Red Sea and Indian Ocean. *Journal of Biogeography*, **40**, 1170–
441 1181.

442 DiBattista J.D., Choat J.H., Gaither M.R., Hobbs J.P., Lozano-Cortés, D.F., Myers R., Paulay G.,
443 Rocha L.A., Toonen R.J., Westneat M., & Berumen M.L. (in review) On the origin of
444 endemic species in the Red Sea. *Journal of Biogeography*.

445 DiBattista J.D., Roberts M., Bouwmeester J., Bowen B.W., Coker D.F., Lozano-Cortés D.F.,
446 Choat J.H., Gaither M.R., Hobbs J.P., Kahil M., Kochzius M., Myers R., Paulay G.,
447 Robitzch V., Saenz-Agudelo P., Salas E., Sinclair-Taylor T.H., Toonen R.J., Westneat M.,
448 Williams S., & Berumen M.L. (in review) A review of contemporary patterns of endemism
449 for shallow water reef fauna in the Red Sea. *Journal of Biogeography*.

450 DiBattista J.D., Waldrop E., Bowen B.W., Schultz J.K., Gaither M.R., Pyle R.L., & Rocha L.A.
451 (2012) Twisted sister species of pygmy angelfishes: discordance between taxonomy,
452 coloration, and phylogenetics. *Coral Reefs*, **31**, 839–851.

453 DiBattista J.D., Wilcox C., Craig M.T., Rocha L. a., & Bowen B.W. (2011) Phylogeography of
454 the Pacific Blueline Surgeonfish, *Acanthurus nigroris*, reveals high genetic connectivity and
455 a cryptic endemic species in the Hawaiian Archipelago. *Journal of Marine Biology*, **2011**,
456 Article ID 839134: 1–17.

457 Drummond A.J., Suchard M.A., Xie D., & Rambaut A. (2012) Bayesian phylogenetics with
458 BEAUti and the BEAST 1.7. *Molecular Biology and Evolution*, **29**, 1969–1973.

459 Evanno G., Regnaut S., & Goudet J. (2005) Detecting the number of clusters of individuals using
460 the software STRUCTURE: a simulation study. *Molecular Ecology*, **14**, 2611–2620.

461 Excoffier L. & Lischer H.E.L. (2010) Arlequin suite ver 3.5: a new series of programs to
462 perform population genetics analyses under Linux and Windows. *Molecular Ecology*
463 *Resources*, **10**, 564–567.

464 Excoffier L., Smouse P.E., & Quattro J.M. (1992) Analysis of molecular variance inferred from
465 metric distances among DNA haplotypes: Application to human mitochondrial DNA
466 restriction data. *Genetics*, **131**, 479–491.

467 Fenton M., Geiselhart S., Rohling E., & Hembelen C. (2000) Aplanktonic zones in the Red Sea.
468 *Marine Micropaleonthology*, **40**, 277–294.

469 Fernandez-Silva I., Moreno E., Essafi A., Fergany M., Garcia-Mas J., Martín-Hernandez A.M.,
470 Alvarez J.M., & Monforte A.J. (2010) Shaping melons: agronomic and genetic
471 characterization of QTLs that modify melon fruit morphology. *Theoretical and Applied*
472 *Genetics*, **121**, 931–940.

Fernandez-Silva I., Snelgrove B.N., & Bowen B.W. (2013) Twelve microsatellite DNA markers to resolve population structure of the Yellow-Striped Goatfish *Mulloidichthys flavolineatus* (family Mullidae). *Conservation Genetics Resources*, **5**, 565–568.

Gaither M.R., Jones S.A., Kelley C., Newman S.J., Sorenson L., & Bowen B.W. (2011) High connectivity in the deepwater snapper *Pristipomoides filamentosus* (Lutjanidae) across the Indo-Pacific with isolation of the Hawaiian archipelago. *PloS one*, **6**, e28913.

Gaither M.R. & Rocha L.A. (2013) Origins of species richness in the Indo-Malay-Philippine biodiversity hotspot: evidence for the centre of overlap hypothesis. *Journal of Biogeography*, **40**, 1638–1648.

Golani D. (2001) *Upeneus davidaromi*, a new deep water goatfish (Osteichthyes, Mullidae) from the Red Sea. *Israel Journal of Zoology*, **47**, 111–121.

Gosline W. (1955) The inshore fish fauna of Johnston Island, a central Pacific atoll. *Pacific Science*, **9**, 442–480.

Guindon S. & Gascuel O. (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology*, **52**, 696–704.

Hedrick P. (2005) A standardized genetic differentiation measure. *Evolution*, **59**, 1633–1638.

Hodge J.R., van Herwerden L., & Bellwood D.R. (2014) Temporal evolution of coral reef fishes: global patterns and disparity in isolated locations. *Journal of Biogeography*, **41**, 2115–2127.

Hodge J.R., Read C.I., van Herwerden L., & Bellwood D.R. (2012) The role of peripheral endemism in species diversification: evidence from the coral reef fish genus *Anampses* (Family: Labridae). *Molecular Phylogenetics and Evolution*, **62**, 653–663.

Hourigan T.F. & Reese E.S. (1987) Mid-ocean isolation and the evolution of Hawaiian reef fishes. *Trends in Ecology & Evolution*, **2**, 187–191.

Jost L. (2008) GST and its relatives do not measure differentiation. *Molecular Ecology*, **17**, 4015–4026.

Kemp J. (1998) Zoogeography of the coral reef fishes of the Socotra Archipelago. *Journal of Biogeography*, **25**, 919–933.

Kimura M. (1980) A simple method for estimating evolutionary rates of base substitution through comparative studies of nucleotide sequences. *Journal of Molecular Evolution*, **16**, 111–120.

Kobayashi D.R. (2006) Colonization of the Hawaiian Archipelago via Johnston Atoll: a characterization of oceanographic transport corridors for pelagic larvae using computer simulation. *Coral Reefs*, **25**, 407–417.

Kulbicki M., Parravicini V., Bellwood D.R., Arias-Gonzàlez E., Chabanet P., Floeter S.R., Friedlander A., McPherson J., Myers R.E., Vigliola L., & Mouillot D. (2013) Global biogeography of reef fishes: a hierarchical quantitative delineation of regions. *PLoS One,* **8**, e81847.

Leis J. & Carson-Ewart B. (1998) Complex behaviour by coral-reef fish larvae in open-water and near-reef pelagic environments. *Environmental Biology of Fishes*, **53**, 259–266.

Leis J.M. (2006) Are larvae of demersal fishes plankton or nekton? *Advances in Marine Biology*, **51**, 57–141.

Lessios H.A. (2008) The Great American Schism: Divergence of marine organisms after the rise of the Central American isthmus. *Annual Review of Ecology, Evolution, and Systematics*, **39**, 63–91.

Lessios H.A. & Robertson D.R. (2013) Speciation on a round planet: phylogeography of the goatfish genus *Mulloidichthys*. *Journal of Biogeography*, **40**, 2373–2384.

Lowe W.H. & Allendorf F.W. (2010) What can genetics tell us about population connectivity? *Molecular Ecology*, **19**, 3038–3051.

Luiz O.J., Allen A.P., Robertson D.R., Floeter S.R., Kulbicki M., Vigliola L., Becheler R., & Madin J.S. (2013) Adult and larval traits as determinants of geographic range size among tropical reef fishes. *Proceedings of the National Academy of Sciences of the United States of America*, **110**, 16498–16502.

Meeker N.D., Hutchinson S.A., Ho L., & Trede N.S. (2007) Method for isolation of PCR-ready genomic DNA from zebrafish tissues. *BioTechniques* **43**, 4–6.

Meirmans P.G. (2012) The trouble with isolation by distance. *Molecular Ecology*, **21**, 2839–2846.

Meirmans P.G. & Van Tienderen P.H. (2004) Genotype and Genodive: two programs for the analysis of genetic diversity of asexual organisms. *Molecular Ecology Notes*, **4**, 792–794.

Meyer A. (1993) Evolution of mitochondrial DNA in fishes. *Biochemistry and molecular biology of fishes*, **2**, 1–38.

Narum S.R. (2006) Beyond Bonferroni: Less conservative analyses for conservation genetics. *Conservation Genetics*, **7**, 783–787.

Nei M. (1987) *Molecular Evolutionary Genetics.* Columbia University Press, New York.

Nei M. & Li W. (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences of the United States of America*, **76**, 5269–5273.

539 Nuryanto A. & Kochzius M. (2009) Highly restricted gene flow and deep evolutionary lineages
540 in the giant clam *Tridacna maxima*. *Coral Reefs*, **28**, 607–619.

541 Planes S. & Fauvelot C. (2002) Isolation by distance and vicariance drive genetic structure of a
542 coral reef fish in the Pacific Ocean. *Evolution*, **56**, 378–399.

543 Pritchard J.K., Stephens M., & Donnelly P. (2000) Inference of population structure using
544 multilocus genotype data. *Genetics*, **155**, 945–959.

545 Pyle R.L. (1999) Patterns of coral reef fish biogeography in the Pacific region. *Marine and*
546 *Coastal Biodiversity in the Tropical Island Pacific Region. Volume 2. Population,*
547 *Development, and Conservation Priorities.* (ed. by and H.F.T. (Eds.). Eldredge, L.G., J.E.
548 Maragos, P.F. Holthus), pp. 157–175. East-West Center/Pacific Science Association,
549 Bishop Museum, Honolulu, Hawaii.

550 Rampen S.W., Schouten S., Koning E., Brummer G.-J.A., & Damsté J.S.S. (2008) A 90 kyr
551 upwelling record from the northwestern Indian Ocean using a novel long-chain diol index.
552 *Earth and Planetary Science Letters*, **276**, 207–213.

553 Randall J. (2007) *Reef and shore fishes of the Hawaiian Islands.* Sea Grant College Program,
554 University of Hawai‘i, Honolulu, Hawaii.

555 Randall J.E. (2002) Mullidae, goatfishes (surmullets). *The Living Marine Resources of the*
556 *Western Central Pacific, Bony fishes part 3 (Menidae to Pomacentridae)* (ed. by K.E.
557 Carpenter and V.H. Niem), pp. 3175–3200. FAO, Rome.

558 Randall J.E., Bogorodsky S., & Krupp F. *Coastal Fishes of the Red Sea.* Fauna of Arabia 26, in
559 press.

560 Rocha L.A, Robertson D.R., Roman J., & Bowen B.W. (2005) Ecological speciation in tropical
561 reef fishes. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, **272**,
562 573–579.

563 Rocha L.A., Craig M.T., & Bowen B.W. (2007) Phylogeography and the conservation of coral
564 reef fishes. *Coral Reefs*, **26**, 501–512.

565 Rogers A. & Harpending H. (1992) Population growth makes waves in the distribution of
566 pairwise genetic differences. *Molecular Biology and Evolution*, **9**, 552–569.

567 Rousset F. (1997) Genetic differentiation and estimation of gene flow from a F-statistics under
568 isolation with distance. *Genetics*, **145**, 1219–1228.

569 Schneider S. & Excoffier L. (1999) Estimation of past demographic parameters from the
570 distribution of pairwise differences when the mutation rates vary among sites: application to
571 human mitochondrial. *Genetics*, **152**, 1079–1089.

572 Siddall M., Rohling E.J., Almogi-Labin A., Hemleben C., Meishner D., Schmelzer I., & Smeed
573 D.A. (2003) Sea-level fluctuations during the last glacial cycle. *Nature*, **423**, 853-858

574 Tamura K., Stecher G., Peterson D., Filipski A., & Kumar S. (2013) MEGA6: Molecular
575 Evolutionary Genetics Analysis version 6.0. *Molecular Biology and Evolution*, **30**, 2725–
576 2729.

577 Tikochinski Y., Shainin I., Hyams Y., Motro U., & Golani D. (2013) Genetic evidence for an
578 undescribed species previously considered as *Sillago sihama* from the northern Red Sea.
579 *Marine Biology Research*, **9**, 309–315.

580 Weir B. & Cockerham C. (1984) Estimating F-statistics for the analysis of population structure.
581 *Evolution*, **6**, 1358–1370.

582 Willette, D.A., Allendorf, F.W., Barber, P.H., Barshis, D.J., Carpenter, K.E., Crandall,
583 E.D., Cresko, W.A., Fernandez-Silva, I., Matz, M.V., Meyer, E., Santos, M.D., Seeb,
584 L.W., Seeb, J.E. (2014) So, you want to use next-generation sequencing in marine systems?
585 Insight from the Pan-Pacific Advanced Studies Institute. *Bulletin of Marine Science* 90, 79-
586 122.

587

588 **SUPPORTING INFORMATION**

589 Additional Supporting Information may be found in the online version of this article:

590

591 **Appendix 1** Supplementary methods.
592 **Appendix 2** Supplementary Figure S1 and Tables S1-S5.
593 **Appendix 3** Supplementary Table S6.

594

595

596

597

598

599 ## Biosketch

600 The authors' research is focused on the evolutionary processes that generate marine biodiversity

601 with the ultimate goal of informing marine conservation. They conduct phylogeographical

surveys of marine species in the tropical Indo-Pacific to test models of speciation, hybridization and dispersal.

Author contributions:

I.F.S., B.W.B., and L.A.R. conceived the project; I.F.S. and J.D.D. assembled the tissue collection with contributions from all the authors; I.F.S. produced and analysed the genetic data, with contributions from R.R.C.; J.E.R. conducted the morphological analyses; B.W.B., L.A.R. and J.D.R. provided lab space and funding support; I.F.S. led the writing, with contributions from all the authors.

Editor: Gustav Paulay

Table 1. Molecular diversity indices for populations of *Mulliodichthys flavolineatus* sampled at sites throughout the Red Sea, the Arabian Sea, the Indian Ocean and the Pacific Ocean. Sample location, number of individuals (*N*), average number of alleles per locus (*N*a), the multi-locus average effective number of alleles (*eff_allele*), observed heterozygosity (H_O), expected heterozygosity (H_E) in the microsatellite dataset are listed by location; number of individuals (*N*), number of haplotypes (N_h), haplotype diversity (*h*) and nucleotide diversity (π) in cytochrome *b* (cyt *b*) dataset are listed by location. Fu's F_S index with associated *P*-values are also given, negative and significant values of Fu's F_S ($P < 0.02$) are indicated in bold.

Sites				cyt *b*				Microsatellites				
	Ocean/Sea	**Lat.**	**Long.**	***N***	***N_h***	***H***	***π***	***N***	***N_a***	***eff_allele***	***H_O***	***H_E***
Greater Indo-Pacific Region												
Big Island (Hawai'i)	Pacific	19.71	-155.45	44	9	0.52	0.002	47	5.92	3.02	0.58	0.61
Johnston Atoll	Pacific	16.73	-169.53	21	8	0.68	0.003	21	5.08	2.79	0.53	0.58
Palmyra (Line Islands)	Pacific	5.87	-162.08	34	10	0.83	0.002	18	6.25	3.82	0.68	0.70
Rarotonga (Cook Islands)	Pacific	-21.23	-159.78	10	8	0.96	0.004	13	5.75	4.15	0.65	0.73
Yoron (Ryukyus)	Pacific	27.05	128.43	16	6	0.68	0.002	24	6.67	3.80	0.62	0.68
Mo'orea (Society Islands)	Pacific	-17.53	-149.83	4	4	1.00	0.007	-	-	-	-	-
Saipan (Marianas)	Pacific	15.18	145.75	2	2	1.00	0.004	-	-	-	-	-
Madagascar	Indian	-25.32	46.17	17	9	0.88	0.006	19	6.42	3.65	0.57	0.68
Mahé (Seychelles)	Indian	-4.68	55.48	4	2	0.67	0.001	-	-	-	-	-
Diego Garcia (Chagos)	Indian	-7.31	72.41	1	1	-	-	-	-	-	-	-
Faafu Atoll (Maldives)	Indian	3.06	72.88	2	1	-	-	-	-	-	-	-
North-western Indian Province												
Muscat (Oman)	Arabian	23.79	58.59	1	1	-	-	-	-	-	-	-
Djibouti	Arabian	11.75	43.21	6	4	0.87	0.005	7	4.67	3.30	0.63	0.64
Farasan Banks	Red	16.75	41.49	5	5	1.00	0.004	-	-	-	-	-
Sudan	Red	21.02	37.13	10	5	0.76	0.005	10	5.58	3.70	0.57	0.65
Jeddah	Red	21.49	39.11	23	10	0.83	0.005	26	6.25	3.61	0.61	0.66
Thuwal	Red	22.33	38.85	5	4	0.90	0.005	-	-	-	-	-
Yanbu	Red	24.07	38.02	2	2	1.00	0.004	-	-	-	-	-
Gulf of Aqaba/Eilat	Red	29.53	35.94	10	7	0.87	0.004	-	-	-	-	-

618 **Table 2.** Tests of neutrality and demographic stability for *Mulloidichthys flavolineatus* sampled at sites described in Table 1

619 throughout the North-western Indian and Indo-Pacific regions. Significant values of Fu's F_S ($P < 0.05$), Tajima's D ($P < 0.02$) and

620 Harpending's Raggedness index r ($P < 0.05$) are indicated in bold.

Geographic region	Fu's *F*	*P*-value	Tajima's *D*	*P*-value	Raggedness Index (*r*)	*P*-value
North-western Indian province	**-6.90**	**0.004**	-1.24	0.092	0.015	0.973
Indo-Pacific (all sites included)	**-28.55**	**<0.001**	**-2.32**	**<0.001**	0.057	0.253
Indo-Pacific (excluding Hawai'i & Johnston A.)	**-21.63**	**<0.001**	**-2.18**	**0.001**	**0.086**	**0.040**
Hawaiian province (Hawai'i & Johnston A.)	**-10.85**	**<0.001**	**-1.84**	**0.011**	0.092	0.739

Table 3. Tests of differentiation among population pairs of *Mulloidichthys flavolineatus* sampled at sites throughout the Red Sea, the Arabian Sea, the Indian Ocean and the Pacific Ocean based on mitochondrial cytochrome *b* (cyt *b*) gene sequences (below diagonal) and microsatellites (above diagonal). Mitochondrial cyt *b* pairwise Φ_{ST} estimates are indicated, whereas conventional F_{ST} values are indicated for microsatellites. Bold denotes values that were significant in the tests of population differentiation after false discovery rate correction (corrected $\alpha_{99.0\%} = 0.002$). Multilocus F_{ST} values after removal of locus mflssr034 from the dataset are indicated between brackets (above the diagonal). Only populations with samples sizes $N > 5$ were included in the analyses, and no microsatellite data for the Gulf of Aqaba/Eilat sample were available.

	Hawai'i	**Johnston Atoll**	**Line Is.**	**Cook Is.**	**Ryukyus**	**Madagascar**	**Djibouti**	**Sudan**	**Jeddah**
Big Island (Hawai'i)	0	-0.003 (-0.005)	**0.053** (**0.056**)	**0.038** (**0.043**)	**0.097** (**0.046**)	**0.074** (**0.063**)	**0.187** (**0.191**)	**0.196** (**0.181**)	**0.135** (**0.132**)
Johnston Atoll	-0.002	0	**0.054** (**0.060**)	**0.056** (**0.062**)	**0.102** (**0.050**)	**0.097** (**0.072**)	**0.201** (**0.209**)	**0.214** (**0.196**)	**0.148** (**0.146**)
Palmyra (Line Is.)	0.032	0.030	0	-0.005 (-0.004)	**0.05** (-0.001)	**0.037** (0.005)	**0.100**(**0.103**)	**0.113** (**0.088**)	**0.086** (**0.080**)
Rarotonga (Cook Is.)	0.082	0.057	-0.011	0	**0.05** (0.002)	0.021 (-0.001)	**0.089** (**0.088**)	**0.096** (**0.071**)	**0.077** (**0.072**)
Yoron (Ryukyus)	0.037	0.036	-0.016	-0.007	0	**0.063** (0.010)	**0.154** (**0.117**)	**0.136** (**0.095**)	**0.116** (**0.079**)
Madagascar	0.063	0.058	0.030	0.007	-0.022	0	**0.135** (**0.100**)	**0.111** (**0.076**)	**0.096** (**0.074**)
Djibouti	**0.923**	**0.900**	**0.896**	**0.866**	**0.902**	**0.856**	0	0.015 (-0.005)	0.006 (0.004)
Sudan	**0.921**	**0.899**	**0.896**	**0.872**	**0.900**	**0.863**	-0.045	0	0.001 (-0.003)
Jeddah	**0.820**	**0.766**	**0.793**	**0.723**	**0.752**	**0.747**	-0.026	-0.031	0
Gulf of Aqaba/Eilat	**0.913**	**0.886**	**0.887**	**0.853**	**0.884**	**0.851**	-0.033	-0.069	-0.026

Corrected values of alpha (after false discovery rate correction)
$\alpha_{95.0\%}= 0.01$, $\alpha_{99.0\%}= 0.002$, $\alpha_{99.9\%}=0.0002$

Table 4. Genetic structuring (Analysis of Molecular Variance, AMOVA) of *Mulloidichthys flavolineatus* sampled at sites throughout the North-western Indian province (Madagascar, Djibouti, Sudan and Saudi Arabia) and the greater Indo-Pacific region (Hawai'i, Johnston Atoll., Line Islands., Cook Islands., Ryukyus and Madagascar) indicated in table 1 based on 715 bp of mtDNA cyt b sequence and microsatellite data sampled at 12 loci. Analyses were conducted 1) with all populations divided into two (North-western Indian province and remaining Indo-Pacific) or three separate groups (North-western Indian province, Hawaiian province and remaining Indo-Pacific) to assess the relationship between these regions; 2) Excluding populations in the North-western Indian province. Φ_{CT}: region variance component relative to total variance; Φ_{SC}: between population within region variance component divided by the sum of itself and within population variance; Φ_{ST}: population variance component relative to the total variance

		Among groups			**Among populations within groups**			**Among populations**	
		% Variation	**Φ_{CT}**	***P* value**	**% Variation**	**Φ_{SC}**	***P* value**	**Φ_{ST}**	***P* value**
All populations included:									
NW Indian vs. Indo-Pacific (including Hawai'i+Johnston A.)	*mtDNA*	90.83	**0.91**	**0.01**	0.00	0.00	0.55	0.91	<0.001
	Msats (all loci)	9.66	**0.10**	**0.01**	4.29	0.05	<0.001	0.14	<0.001
Hawai'i vs. rest Indo-Pacific vs. NW Indian	*mtDNA*	84.01	**0.84**	**0.01**	-0.12	-0.01	0.78	0.84	<0.001
	Msats (all loci)	8.46	**0.08**	**0.001**	2.20	0.02	<0.001	0.11	<0.001
NW Indian vs. Madagascar vs. (Cook Is.+Line Is.+Ryukyus)	*mtDNA*	72.24	**0.72**	**0.006**	-0.25	-0.01	0.90	0.72	<0.001
vs.(Hawai'i+Johnston A.)	*Msats (all loci)*	8.17	**0.08**	**0.001**	-0.39	0.00	0.96	0.08	<0.001
North-western Indian province excluded:									
(Hawai'i+Johnston A.) vs. Cook Is. vs. (Line Is.+Ryukyus+Madagascar)	*mtDNA*	2.99	0.03	0.07	-0.17	0.00	0.53	0.03	0.07
	Msats (excluding ssr037)	4.93	**0.05**	**0.04**	0.22	0.00	0.41	0.05	<0.001
(Hawai'i+Johnston A.) vs. Cook Is. vs. Line Is.	*mtDNA*	4.82	**0.05**	**0.02**	-2.15	-0.02	0.88	0.03	0.07
vs. (Ryukyus+Madagascar)	*Msats (excluding ssr037)*	4.05	0.04	0.13	0.35	0.00	0.25	0.04	<0.001
(Hawai'i+Johnston A.) vs. (Cook Is.+Line Is.+Ryukyus) vs. Madagascar	*mtDNA*	3.86	0.04	0.07	-0.01	-0.01	0.60	0.03	0.07
	Msats (excluding ssr037)	3.88	0.04	0.07	0.02	0.00	0.63	0.04	<0.001
(Hawai'i+Johnston A.) vs. (Cook Is.+Line Is.+Ryukyus+Madagascar)	*mtDNA*	95.63	0.04	0.07	0.62	0.00	0.38	0.04	0.01
	Msats (excluding ssr037)	5.17	0.05	0.07	-0.16	0.00	0.85	0.05	<0.001

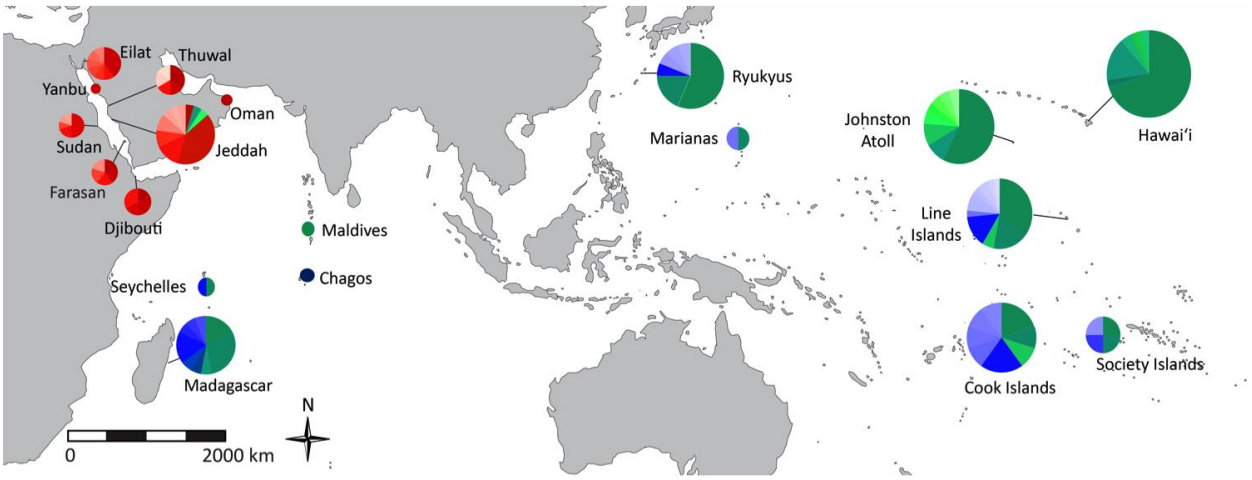
Titles and legends to figures

Figure 1. Map of the study area. Pie charts represent the proportion of haplotypes sampled at each site (one pie chart per sampling site with the size proportional to sample size). Across sampling sites, each colour represents a unique haplotype. Haplotypes of the North-western Indian lineage are represented in red, haplotypes of the Indo-Pacific lineage are represented in green and blue (as defined in Fig. 2).

Figure 2. Median-joining network based on 715 base pairs of mitochondrial cytochrome *b* sequence data (N = 217) from *Mulloidichthys flavolineatus* sampled across the Red Sea, Arabian Sea, Indian Ocean and Pacific Ocean. Each circle represents a haplotype, and its size is proportional to its total frequency. Branches separated by black crossbars represent a single nucleotide change, whereas open circles indicate unsampled haplotypes; colours denote collection location as indicated by the embedded key. The network depicts two distinct clades separated by 7 mutational steps (corrected sequence divergence, d = 1.7%; Kimura, 1980). The photo on the top, taken in Hawai‘i, depicts specimens of the Indo-Pacific lineage with a white/grey caudal fin. On the bottom is a photo of goatfishes of the North-western Indian lineage with yellow caudal fins, taken on Thuwal reefs in the Saudi Arabian Red Sea. (Photo credits: M. Royer and T. Sinclair-Taylor).

Figure 3. Assignment probabilities of *Mulloidichthys flavolineatus* individuals to K = 3 clusters based on 12 microsatellite loci, using the program STRUCTURE, with individuals arranged geographically (west to east) along the x-axis.

Proposed caption for cover image for **Yellow tails in a Red Sea: Phylogeography of the Indo-Pacific goatfish *Mulloidichthys flavolineatus* reveals isolation in peripheral provinces and cryptic evolutionary lineages,** by Iria Fernandez-Silva, John E. Randall, Richard R. Coleman, Joseph D. DiBattista, Luiz A. Rocha, James D. Reimer, Carl G. Meyer, Brian W. Bowen, reads "Goatfishes of *Mulloidichthys flavolineatus* with yellow tails in the NW Indian Oceaa (T. Sinclair-Taylor)"



677

678

679

680

681

682

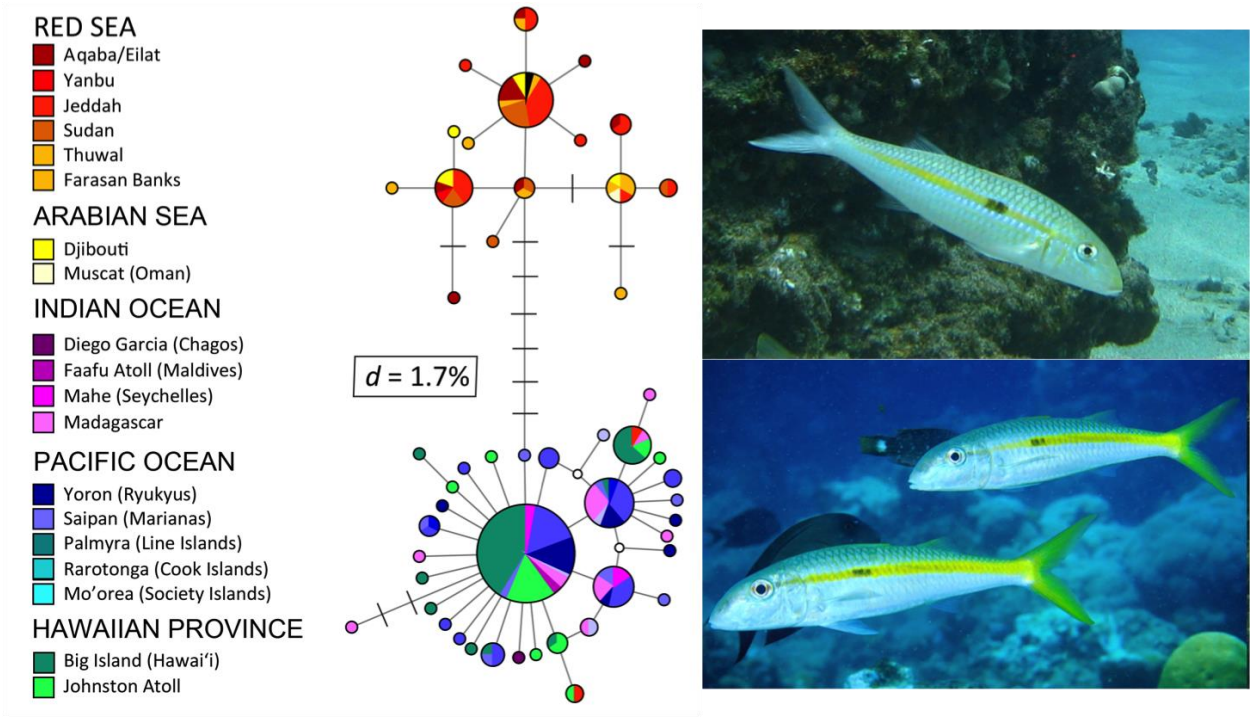
683

684

685

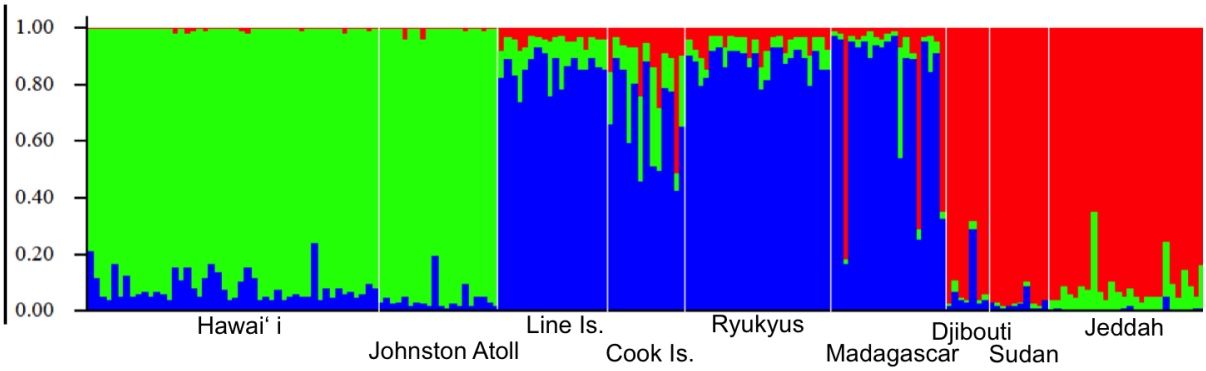
686

687


RED SEA
Aqaba/Eilat
Yanbu
Jeddah
Sudan
Thuwal
Farasan Banks
ARABIAN SEA
Djibouti
Muscat (Oman)
INDIAN OCEAN
Diego Garcia (Chagos)
Faafu Atoll (Maldives)
Mahe (Seychelles)
Madagascar
PACIFIC OCEAN
Yoron (Ryukyus)
Saipan (Marianas)
Palmyra (Line Islands)
Rarotonga (Cook Islands)
Mo'orea (Society Islands)
HAWAIIAN PROVINCE
Big Island (Hawai'i)
Johnston Atoll
d = 1.7%

688

689


1.00
0.80
0.60
0.40
0.20
0.00
Hawai‘ i
Johnston Atoll
Line Is.
Cook Is.
Ryukyus
Madagascar
Djibouti
Sudan
Jeddah

690

691